1636

#17 1600

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/689,343D

DATE: 10/22/2002
TIME: 16:29:38

Input Set : A:\NEB-181-2.txt
Output Set: N:\CRF4\10222002\I689343D.raw

<110> APPLICANT: Vaisvila, Romualdus
Morgan, Richard D.
Kucera, Rebecca B.
Claus, Toby B.
Raleigh, Elisabeth A.

<120> TITLE OF INVENTION: Method For Cloning And Producing The MseI Restriction Endonuclease

<130> FILE REFERENCE: NEB-181

<140> CURRENT APPLICATION NUMBER: US 09/689,343D

<141> CURRENT FILING DATE: 2000-10-12

<160> NUMBER OF SEQ ID NOS: 9

<170> SOFTWARE: PatentIn version 3.1

<210> SEQ ID NO: 1

<211> LENGTH: 903

<212> TYPE: DNA

<213> ORGANISM: Micrococcus sp.

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (1)..(900)

<223> OTHER INFORMATION:

W--> <400> 1

```
atg cct atc tcg acc gtc tgg acg ccg gac gga gac gac ctc atc gtg       48
Met Pro Ile Ser Thr Val Trp Thr Pro Asp Gly Asp Asp Leu Ile Val
1               5                  10                  15

gag gcg gac aac ctc gat ttc att caa acg ctc ccc gac gcg agc ttc       96
Glu Ala Asp Asn Leu Asp Phe Ile Gln Thr Leu Pro Asp Ala Ser Phe
            20                  25                  30

cga atg atc tac atc gat ccg ccg ttc aac aca ggg cga acg cag cgg      144
Arg Met Ile Tyr Ile Asp Pro Pro Phe Asn Thr Gly Arg Thr Gln Arg
        35                  40                  45

ctt cag tcg ctc aag acg acc cgc tcg gtc aca ggg tcg cga gtc ggc      192
Leu Gln Ser Leu Lys Thr Thr Arg Ser Val Thr Gly Ser Arg Val Gly
    50                  55                  60

ttc aaa ggc cag acg tac gac acg gtc aag agc act ctg cac tcg tat      240
Phe Lys Gly Gln Thr Tyr Asp Thr Val Lys Ser Thr Leu His Ser Tyr
65                  70                  75                  80

gac gac gct ttc acc gac tat tgg tcg ttc ctc gaa ccg cgt ctc ctg      288
Asp Asp Ala Phe Thr Asp Tyr Trp Ser Phe Leu Glu Pro Arg Leu Leu
                85                  90                  95

gag gct tgg cgg ttg ctc acc cct gac ggc gcg ctc tat ctt cat ctg      336
Glu Ala Trp Arg Leu Leu Thr Pro Asp Gly Ala Leu Tyr Leu His Leu
            100                 105                 110

gat tac cgc gag gtt cac tac gcc aag gtc gtc ctc gac gcg atg ttc      384
```

RAW SEQUENCE LISTING DATE: 10/22/2002
PATENT APPLICATION: US/09/689,343D TIME: 16:29:38

Input Set : A:\NEB-181-2.txt
Output Set: N:\CRF4\10222002\I689343D.raw

```
64 Asp Tyr Arg Glu Val His Tyr Ala Lys Val Val Leu Asp Ala Met Phe
65         115                 120                 125
67 gga cgc gaa agc ttc ctg aac gag ctg atc tgg gcg tac gac tac ggc      432
68 Gly Arg Glu Ser Phe Leu Asn Glu Leu Ile Trp Ala Tyr Asp Tyr Gly
69     130                 135                 140
71 gcg cgc tcg aag agc aag tgg ccc acc aag cac gac aac atc ctc gtg      480
72 Ala Arg Ser Lys Ser Lys Trp Pro Thr Lys His Asp Asn Ile Leu Val
73 145                 150                 155                 160
75 tat gtg aag gac ccg aac aac tac gtc tgg aac ggt cag gat gta gat      528
76 Tyr Val Lys Asp Pro Asn Asn Tyr Val Trp Asn Gly Gln Asp Val Asp
77                 165                 170                 175
79 cgc gag ccc tac atg gcg ccc ggg ctc gtt aca ccc gag aag gta gcg      576
80 Arg Glu Pro Tyr Met Ala Pro Gly Leu Val Thr Pro Glu Lys Val Ala
81             180                 185                 190
83 ctt ggc aag ctg ccc acc gac gtc tgg tgg cac aca atc gtt ccg cct      624
84 Leu Gly Lys Leu Pro Thr Asp Val Trp Trp His Thr Ile Val Pro Pro
85         195                 200                 205
87 gcg agc aaa gag cgc acc ggg tac gcg aca cag aag ccg gtc ggc atc      672
88 Ala Ser Lys Glu Arg Thr Gly Tyr Ala Thr Gln Lys Pro Val Gly Ile
89     210                 215                 220
91 atc cgt cgc atg att cag gcg agc agc aat gaa ggc gac tgg gtt ctg      720
92 Ile Arg Arg Met Ile Gln Ala Ser Ser Asn Glu Gly Asp Trp Val Leu
93 225                 230                 235                 240
95 gat ttc ttc gct ggt agt ggg acg acc ggc gcc gcg gcc cgc cag ctc      768
96 Asp Phe Phe Ala Gly Ser Gly Thr Thr Gly Ala Ala Ala Arg Gln Leu
97                 245                 250                 255
99 gga cgc cgt ttt gtg ctc gta gac gtc aac cca gaa gca atc gcg gta      816
100 Gly Arg Arg Phe Val Leu Val Asp Val Asn Pro Glu Ala Ile Ala Val
101             260                 265                 270
103 atg gca aaa cgg ttg gat gac ggg gca ttg gac acc agc gtg acg atc      864
104 Met Ala Lys Arg Leu Asp Asp Gly Ala Leu Asp Thr Ser Val Thr Ile
105         275                 280                 285
107 gtg cag act ccc cag agt gac cca cga acc gac gga tga                 903
108 Val Gln Thr Pro Gln Ser Asp Pro Arg Thr Asp Gly
109     290                 295                 300
111 <210> SEQ ID NO: 2
112 <211> LENGTH: 300
113 <212> TYPE: PRT
114 <213> ORGANISM: Micrococcus sp.
116 <400> SEQUENCE: 2
118 Met Pro Ile Ser Thr Val Trp Thr Pro Asp Gly Asp Asp Leu Ile Val
119 1               5                   10                  15
121 Glu Ala Asp Asn Leu Asp Phe Ile Gln Thr Leu Pro Asp Ala Ser Phe
122             20                  25                  30
124 Arg Met Ile Tyr Ile Asp Pro Pro Phe Asn Thr Gly Arg Thr Gln Arg
125         35                  40                  45
127 Leu Gln Ser Leu Lys Thr Thr Arg Ser Val Thr Gly Ser Arg Val Gly
128     50                  55                  60
130 Phe Lys Gly Gln Thr Tyr Asp Thr Val Lys Ser Thr Leu His Ser Tyr
```

RAW SEQUENCE LISTING DATE: 10/22/2002
PATENT APPLICATION: US/09/689,343D TIME: 16:29:38

Input Set : A:\NEB-181-2.txt
Output Set: N:\CRF4\10222002\I689343D.raw

```
131 65                  70                  75                  80
133 Asp Asp Ala Phe Thr Asp Tyr Trp Ser Phe Leu Glu Pro Arg Leu Leu
134                 85                  90                  95
136 Glu Ala Trp Arg Leu Leu Thr Pro Asp Gly Ala Leu Tyr Leu His Leu
137             100                 105                 110
139 Asp Tyr Arg Glu Val His Tyr Ala Lys Val Val Leu Asp Ala Met Phe
140         115                 120                 125
142 Gly Arg Glu Ser Phe Leu Asn Glu Leu Ile Trp Ala Tyr Asp Tyr Gly
143     130                 135                 140
145 Ala Arg Ser Lys Ser Lys Trp Pro Thr Lys His Asp Asn Ile Leu Val
146 145                 150                 155                 160
148 Tyr Val Lys Asp Pro Asn Asn Tyr Val Trp Asn Gly Gln Asp Val Asp
149                 165                 170                 175
151 Arg Glu Pro Tyr Met Ala Pro Gly Leu Val Thr Pro Glu Lys Val Ala
152             180                 185                 190
154 Leu Gly Lys Leu Pro Thr Asp Val Trp Trp His Thr Ile Val Pro Pro
155         195                 200                 205
157 Ala Ser Lys Glu Arg Thr Gly Tyr Ala Thr Gln Lys Pro Val Gly Ile
158     210                 215                 220
160 Ile Arg Arg Met Ile Gln Ala Ser Ser Asn Glu Gly Asp Trp Val Leu
161 225                 230                 235                 240
163 Asp Phe Phe Ala Gly Ser Gly Thr Thr Gly Ala Ala Ala Arg Gln Leu
164                 245                 250                 255
166 Gly Arg Arg Phe Val Leu Val Asp Val Asn Pro Glu Ala Ile Ala Val
167             260                 265                 270
169 Met Ala Lys Arg Leu Asp Asp Gly Ala Leu Asp Thr Ser Val Thr Ile
170         275                 280                 285
172 Val Gln Thr Pro Gln Ser Asp Pro Arg Thr Asp Gly
173     290                 295                 300
175 <210> SEQ ID NO: 3
176 <211> LENGTH: 1236
177 <212> TYPE: DNA
178 <213> ORGANISM: Unknown
180 <220> FEATURE:
181 <223> OTHER INFORMATION: Environmental DNA
183 <220> FEATURE:
184 <221> NAME/KEY: CDS
185 <222> LOCATION: (1)..(1233)
186 <223> OTHER INFORMATION:
188 <220> FEATURE:
189 <221> NAME/KEY: misc_feature
190 <222> LOCATION: (198)..(198)
191 <223> OTHER INFORMATION: Xaa = any amino acid
W--> 193 <220>
194 <221> NAME/KEY: misc_feature
195 <222> LOCATION: (594)..(594)
196 <223> OTHER INFORMATION: N= G, A, C or T
198 <400> SEQUENCE: 3
200 atg cct aca ctg gat tgg ccc ggt aaa cag tta agc ttc cca cca gct       48
```

RAW SEQUENCE LISTING DATE: 10/22/2002
PATENT APPLICATION: US/09/689,343D TIME: 16:29:38

Input Set : A:\NEB-181-2.txt
Output Set: N:\CRF4\10222002\I689343D.raw

```
201 Met Pro Thr Leu Asp Trp Pro Gly Lys Gln Leu Ser Phe Pro Pro Ala
202 1               5                   10                  15
204 acc tcc ttg cat ctg gag agt gtg gtc act gag gga gcg gag tca ccg      96
205 Thr Ser Leu His Leu Glu Ser Val Val Thr Glu Gly Ala Glu Ser Pro
206             20                  25                  30
208 cct aat cgt ctg att tgg gcg gac aac ctg ccg cta atg gta gat ttg     144
209 Pro Asn Arg Leu Ile Trp Ala Asp Asn Leu Pro Leu Met Val Asp Leu
210         35                  40                  45
212 ttg gcc gaa tat gaa ggg aaa atc gat ctg atc tac gcc gat ccc cct     192
213 Leu Ala Glu Tyr Glu Gly Lys Ile Asp Leu Ile Tyr Ala Asp Pro Pro
214     50                  55                  60
216 ttt ttt acg gat cgt act tat gcg gcg cga att ggt cat ggg gag gat     240
217 Phe Phe Thr Asp Arg Thr Tyr Ala Ala Arg Ile Gly His Gly Glu Asp
218 65                  70                  75                  80
220 tcg cgt cgt cca caa acc tgg cag ctt gca gaa gga tat acg gac gag     288
221 Ser Arg Arg Pro Gln Thr Trp Gln Leu Ala Glu Gly Tyr Thr Asp Glu
222                 85                  90                  95
224 tgg aag gat tta gat gaa tac ctg gac ttc ctt tat cca cgc ctg gta     336
225 Trp Lys Asp Leu Asp Glu Tyr Leu Asp Phe Leu Tyr Pro Arg Leu Val
226             100                 105                 110
229 ctg atg tat cga ctg ctg gca cca cac gga acg ctc tac ttg cac ctg     384
230 Leu Met Tyr Arg Leu Leu Ala Pro His Gly Thr Leu Tyr Leu His Leu
231         115                 120                 125
233 gac tgg cac gcc aat gcc tac gta cgt gta ctg ctt gat gag atc ttc     432
234 Asp Trp His Ala Asn Ala Tyr Val Arg Val Leu Leu Asp Glu Ile Phe
235     130                 135                 140
237 ggg cga cag cgg ttt ctc aac gag atc gtc tgg atc tat cac ggc ccc     480
238 Gly Arg Gln Arg Phe Leu Asn Glu Ile Val Trp Ile Tyr His Gly Pro
239 145                 150                 155                 160
241 tca gcc atc cga cgc gcc ttc aag cgc aaa cat gat acc atc ttg gtt     528
242 Ser Ala Ile Arg Arg Ala Phe Lys Arg Lys His Asp Thr Ile Leu Val
243                 165                 170                 175
245 tat gtg aaa ggt gaa aac tat aca ttc aat gcg gat gcg gtt cgt caa     576
246 Tyr Val Lys Gly Glu Asn Tyr Thr Phe Asn Ala Asp Ala Val Arg Gln
247             180                 185                 190
W-> 249 cct tac cat ccg agc acn cat aag acc ttc gct tcc tcc ccg aag gcc     624
W-> 250 Pro Tyr His Pro Ser Xaa His Lys Thr Phe Ala Ser Ser Pro Lys Ala
251         195                 200                 205
253 ggc ttt ggt aag gtg ccg gat ctg cag cgc ggc aaa gtg ccc gaa gac     672
254 Gly Phe Gly Lys Val Pro Asp Leu Gln Arg Gly Lys Val Pro Glu Asp
255     210                 215                 220
257 tgg tgg tat ttt ccg gtc gtg gcc cgt cta cac cga gaa cgg agc ggc     720
258 Trp Trp Tyr Phe Pro Val Val Ala Arg Leu His Arg Glu Arg Ser Gly
259 225                 230                 235                 240
261 tat ccg act caa aag cct caa gcc ttg ctg gag cgg atc ctg ctg gcc     768
262 Tyr Pro Thr Gln Lys Pro Gln Ala Leu Leu Glu Arg Ile Leu Leu Ala
263                 245                 250                 255
265 tcc tcg aac gca ggc gat ctg gtg gca gac ttc ttc tgc ggc tca ggg     816
266 Ser Ser Asn Ala Gly Asp Leu Val Ala Asp Phe Phe Cys Gly Ser Gly
```

RAW SEQUENCE LISTING DATE: 10/22/2002
PATENT APPLICATION: US/09/689,343D TIME: 16:29:38

Input Set : A:\NEB-181-2.txt
Output Set: N:\CRF4\10222002\I689343D.raw

```
267             260                 265                 270
269 aca acc gct gtg gtg gca gcc cgt ctg gga cgg cgc ttc ctg gtc aac      864
270 Thr Thr Ala Val Val Ala Ala Arg Leu Gly Arg Arg Phe Leu Val Asn
271         275                 280                 285
273 gat gca agc tgg cgc gcc gtt cat gtg aca cgc aca cgc ttg cta cgc      912
274 Asp Ala Ser Trp Arg Ala Val His Val Thr Arg Thr Arg Leu Leu Arg
275     290                 295                 300
277 gag gga gta agt ttc act ttt gaa cgc cag gaa act ttt act cta cct      960
278 Glu Gly Val Ser Phe Thr Phe Glu Arg Gln Glu Thr Phe Thr Leu Pro
279 305                 310                 315                 320
281 atc cag cca ctt cca cca gat tgg ttg atc atc gcc gag gag cag att     1008
282 Ile Gln Pro Leu Pro Pro Asp Trp Leu Ile Ile Ala Glu Glu Gln Ile
283                 325                 330                 335
286 cgc ctc caa gca ccc ttt ctc gta gat ttt tgg gaa gtg gac gat caa     1056
287 Arg Leu Gln Ala Pro Phe Leu Val Asp Phe Trp Glu Val Asp Asp Gln
288             340                 345                 350
290 tgg gat ggc aaa atc ttc cgc agc cgt cat caa ggc tta cgc tcc cgc     1104
291 Trp Asp Gly Lys Ile Phe Arg Ser Arg His Gln Gly Leu Arg Ser Arg
292         355                 360                 365
294 ctt cag gag cag gcg ccg ctc tct cta cca ttg acc ggg aat gga ctg     1152
295 Leu Gln Glu Gln Ala Pro Leu Ser Leu Pro Leu Thr Gly Asn Gly Leu
296     370                 375                 380
298 ttg tgt gta cgg gta gtg agc cgt gaa ggg gaa tac tat gag ttc aca     1200
299 Leu Cys Val Arg Val Val Ser Arg Glu Gly Glu Tyr Tyr Glu Phe Thr
300 385                 390                 395                 400
302 ggt cga gcc gat agc cct cac ccc gta tcg ttt tga                    1236
303 Gly Arg Ala Asp Ser Pro His Pro Val Ser Phe
304                 405                 410
306 <210> SEQ ID NO: 4
307 <211> LENGTH: 411
308 <212> TYPE: PRT
309 <213> ORGANISM: Unknown
311 <220> FEATURE:
312 <223> OTHER INFORMATION: Environmental DNA
314 <220> FEATURE:
315 <221> NAME/KEY: misc_feature
316 <222> LOCATION: (198)..(198)
317 <223> OTHER INFORMATION: Xaa = any amino acid
319 <400> SEQUENCE: 4
321 Met Pro Thr Leu Asp Trp Pro Gly Lys Gln Leu Ser Phe Pro Pro Ala
322 1               5                   10                  15
324 Thr Ser Leu His Leu Glu Ser Val Val Thr Glu Gly Ala Glu Ser Pro
325             20                  25                  30
327 Pro Asn Arg Leu Ile Trp Ala Asp Asn Leu Pro Leu Met Val Asp Leu
328         35                  40                  45
330 Leu Ala Glu Tyr Glu Gly Lys Ile Asp Leu Ile Tyr Ala Asp Pro Pro
331     50                  55                  60
333 Phe Phe Thr Asp Arg Thr Tyr Ala Ala Arg Ile Gly His Gly Glu Asp
334 65                  70                  75                  80
```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 10/22/2002
PATENT APPLICATION: US/09/689,343D TIME: 16:29:39

Input Set : A:\NEB-181-2.txt
Output Set: N:\CRF4\10222002\I689343D.raw

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:3; N Pos. 594
Seq#:3; Xaa Pos. 198
Seq#:4; Xaa Pos. 198